NIA U01AG058532

Multiscale modeling of the Cocktail Party Problem

Mounya Elhilali, Johns Hopkins University
Shihab Shamma, University of Maryland, College Park

Problem statement

Sounds in everyday life seldom exist in isolation. We are constantly surrounded by a cacophony of sounds that constantly impinge on our ears

☛ **Cocktail Party Problem (CPP)**

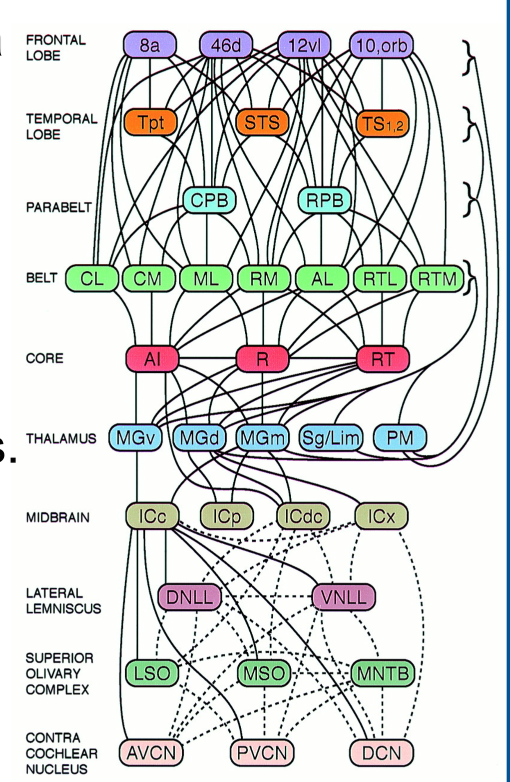
Medical & Engineering relevance:

- ✓ Hearing prostheses
- ✓ Audio technologies
- ✓ Communication systems
- ✓ Medical diagnosis
- ✓ Microphone Design

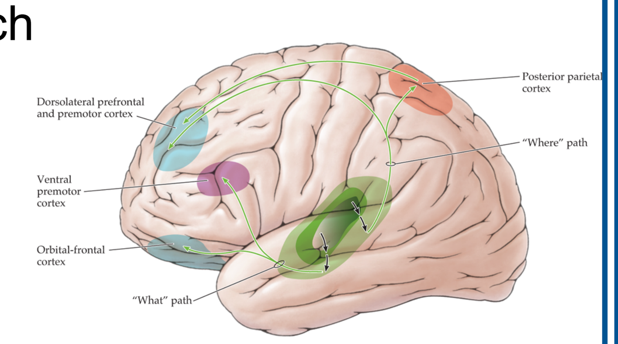
...

Multiscale Approach

- An acoustic signal undergoes a series of transformations from the cochlea all the way to auditory cortex, effectively extracting a rich feature representation that forms the basis for perceptual representation of sound objects.
- These transformations evolve along multiple spectral and temporal scales and engage local and global neural circuits

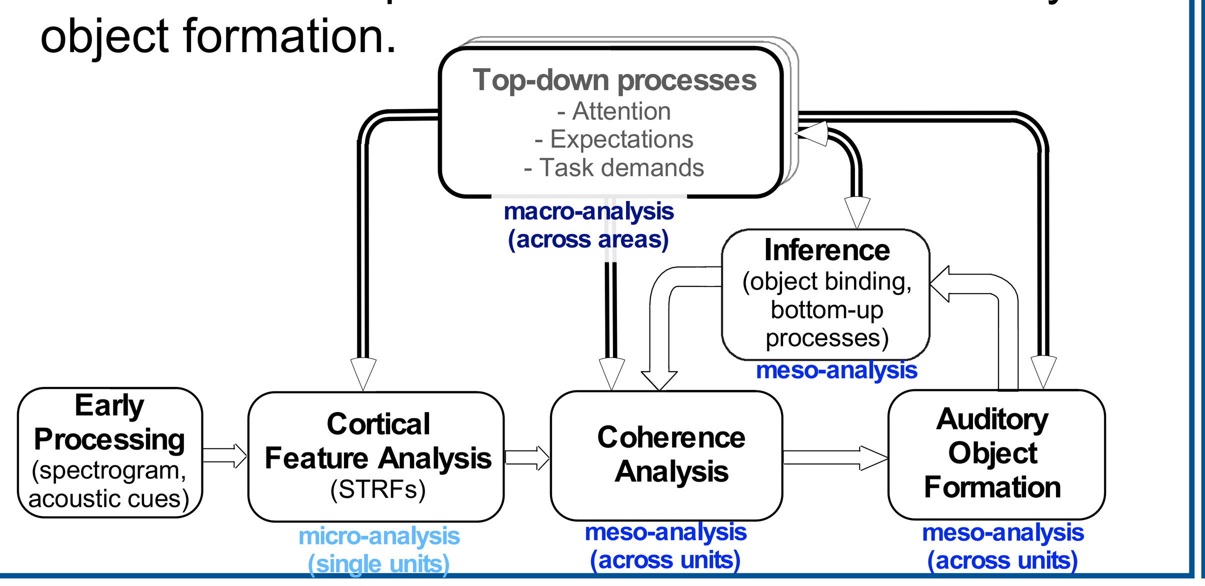


The current project puts forth an **adaptive theory of auditory perception** which integrates the role of *both* sensory mechanisms and cognitive control in a unified multiscale scheme.

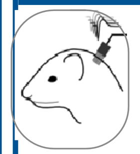


Multiscale modeling:

- *Micro-level* analysis of complex sounds (at the single neuron level) into a multidimensional space
- *Meso-level* group analysis correlating activity in populations of cortical neurons (bottom-up)
- *Macro-level* (across-area) feedback processes of attention and expectations that mediate auditory object formation.

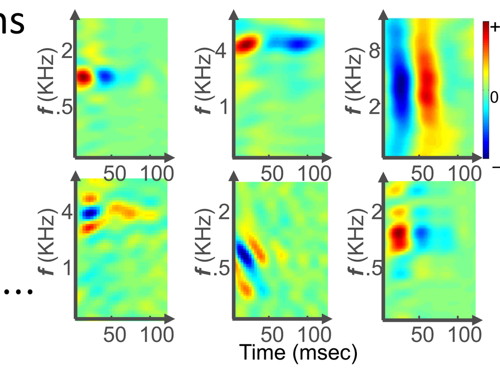


Statistical inference of sound



- **Modeling neural response properties**
  - At level of individual neurons
  - Across neural populations

$$r(t) = \vartheta_0 + \int_{-\infty}^{\infty} \vartheta_1(\tau)s(t-\tau)d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \vartheta_2(\tau_1,\tau_2)s(t-\tau_1)s(t-\tau_2)d\tau_1 d\tau_2 + \cdots$$



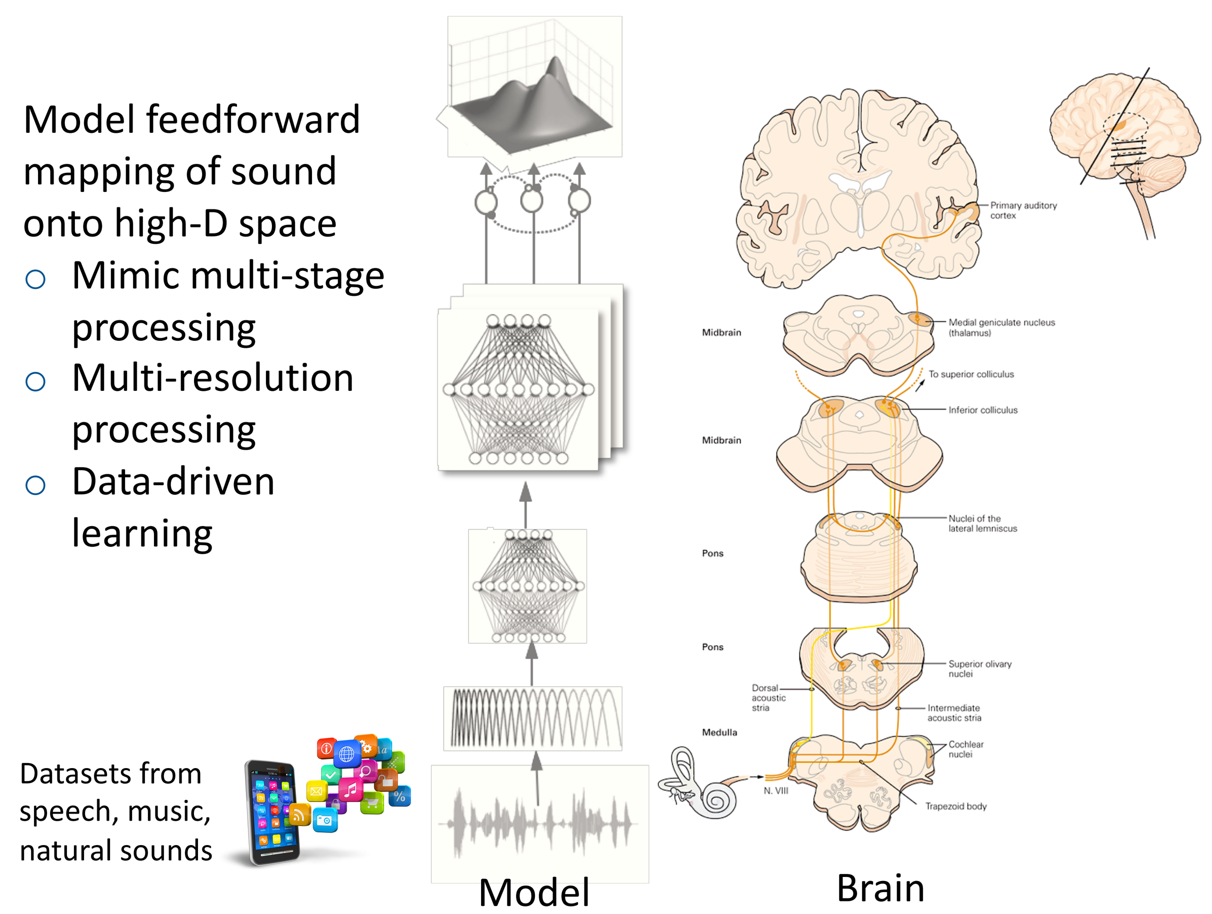
$$\boldsymbol{r(t) \approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \vartheta_2(\tau_1,\tau_2)s(t-\tau_1)s(t-\tau_2)d\tau_1 d\tau_2}$$

Linear approximation using 2nd order Volterra kernel

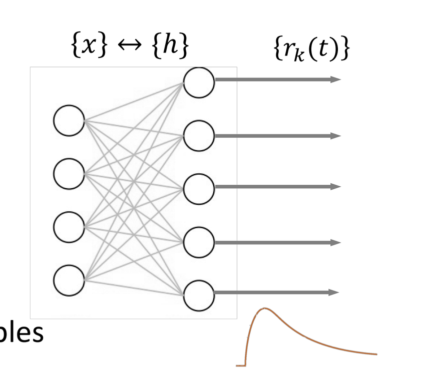
- **Mapping into a high-D statistical space**

Model feedforward mapping of sound onto high-D space

- Mimic multi-stage processing
- Multi-resolution processing
- Data-driven learning



- **Local scale analysis**
  - Sparse **R**estricted **B**oltzmann **M**achine
    Generative stochastic network
    ⇨ **Discover** structure in the data
    Infer **probabilistic distribution** over hidden variables



$$P(x,h) = \frac{1}{Z}e^{-E(x,h)}$$

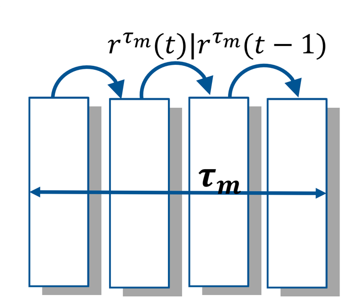
Decaying energy fct.

normalization

Once trained, weight connections are converted into STRFs $\psi(.)$

$$r_k(t) = \sum_f \int S_i(\tau,f).\psi(t-\tau,f)d\tau$$

- **Global scale analysis**
  - Conditional **R**estricted **B**oltzmann **M**achine
    Generative stochastic network
    Incorporates 'priors' from activations over recent past $\boldsymbol{\tau_m}$



$$P(r(t),h(t)|\vec{r}(t)) = \frac{1}{Z}e^{-E\left(r(t),h(t)|\vec{r}(t)\right)}$$

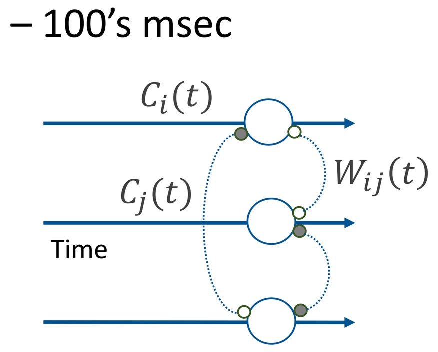
Decaying energy fct.

Sequential integration over 10's – 100's msec

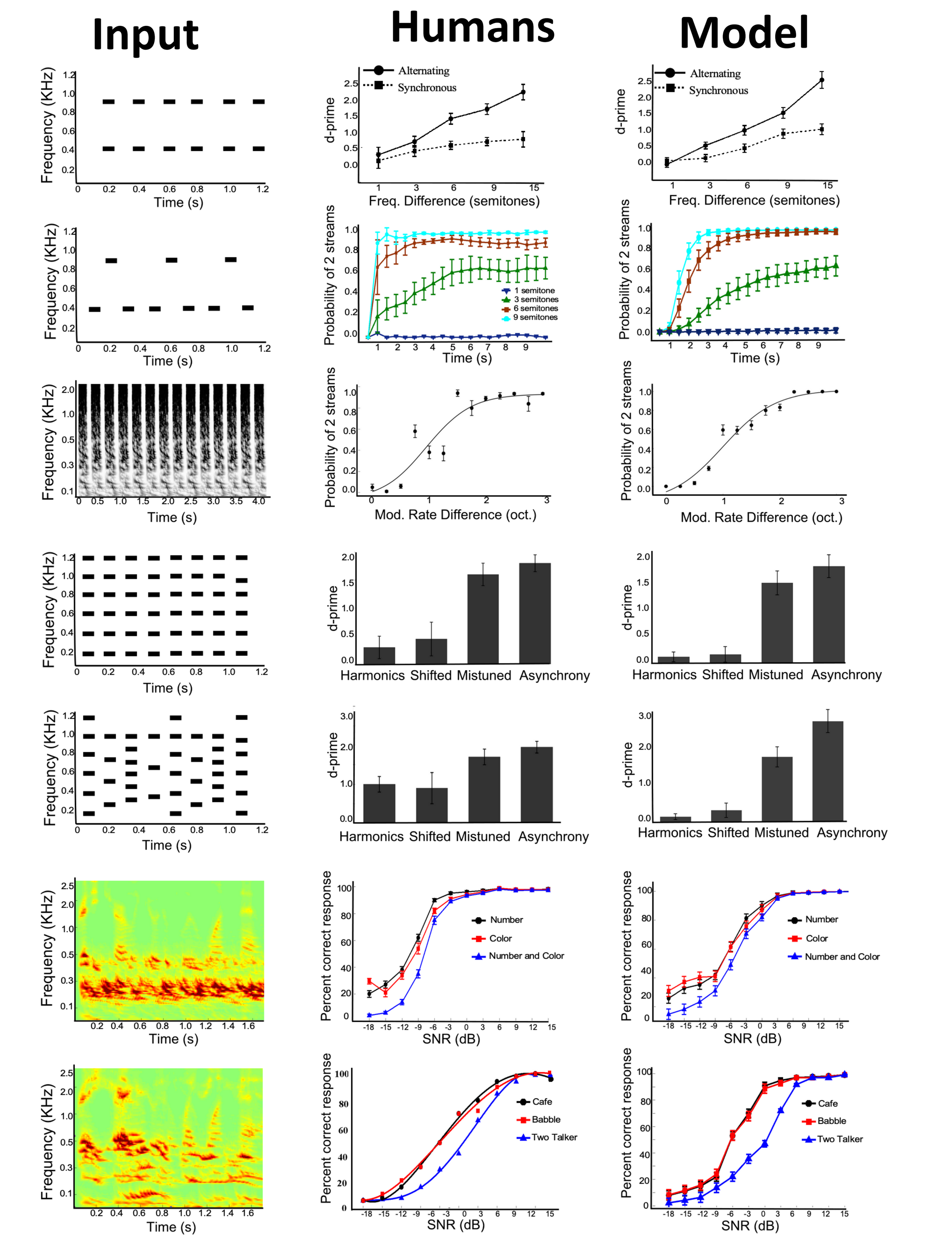
- **Macro-scale integration**

$$W_{ij}(t) = \zeta W_{ij}(t-1) + \xi C_i(t)C_j(t)$$

Hebbian learning



Model Validation

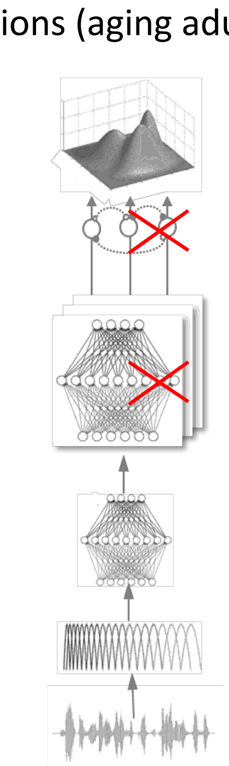


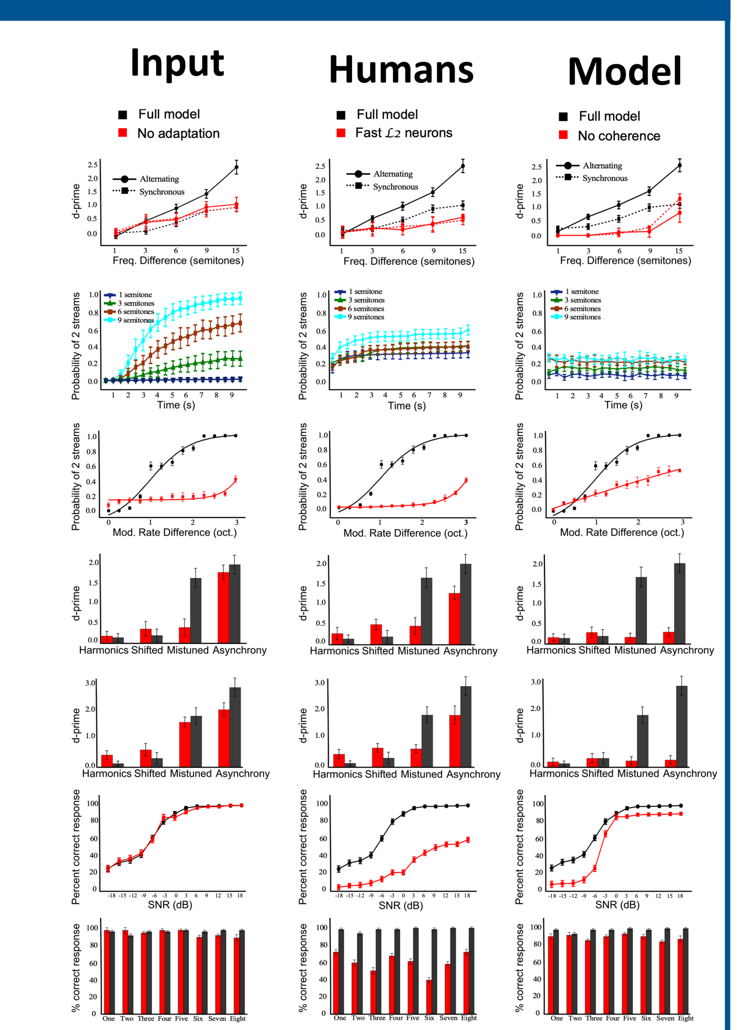
Model Credibility

- **Context:** Multiscale model for auditory stream segregation
- **Data**: Perception of simple and complex sounds by human listeners
- **Evaluation**: Fixed model tested with exact stimuli used for human listeners
- **Limitations**: Model is feedforward with no contextual knowledge (language, memory)
- **Version control** : Implementation history and functionality are documented
- **Dissemination**: Model is publicly shared online, along with peer-reviewed paper
- **Competition**: Model used as backbone for open competition on auditory event detection (DCASE - IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events)
- **Opportunities**: Springboard to expand into multiple neural scales and explore effects of feedback interactions (top-down/bottom-up)

Model function and malfunction

Modify model structure to test limitations, necessary and sufficient processes as well as explore implications for special populations (aging adults)

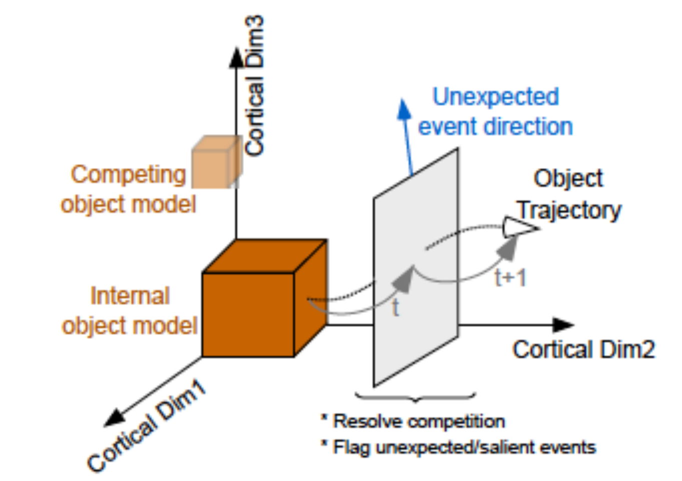




Cognitive control

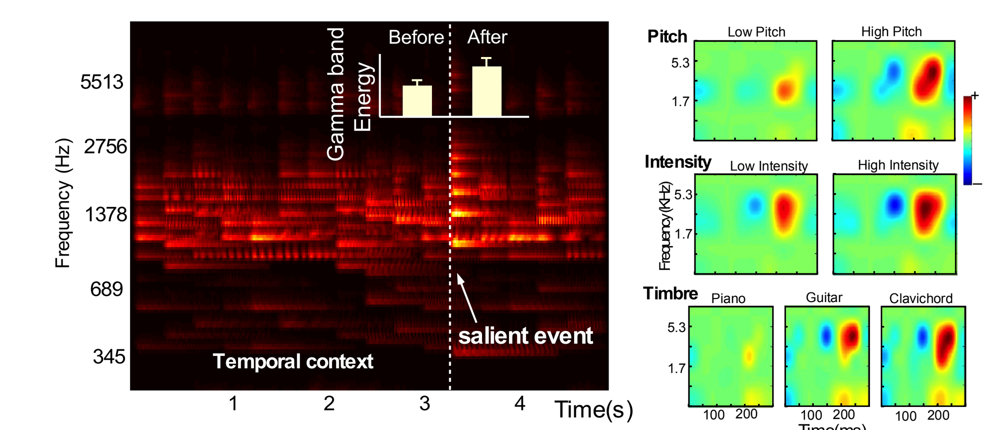
- **Role of working memory and bottom-up attention**

Tracking of sound elements in the cortical high-D space operates as a Bayesian inference process.



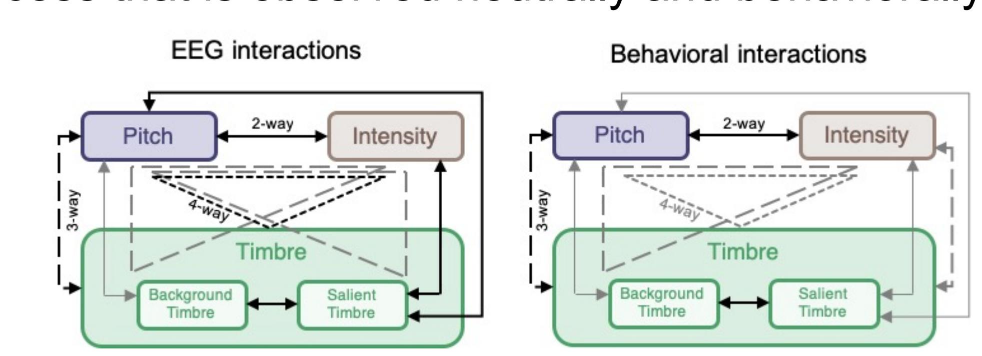
Human listeners tracking dynamical scenes show sensitivity to changes in:

- ✓ Mean and variance changes of acoustic features
- ✓ High-order statistics of individual features



Changes in neural responses (EEG) of dynamic scenes driven by bottom-up attention

Integration across acoustic attributes is a nonlinear process that is observed neutrally and behaviorally



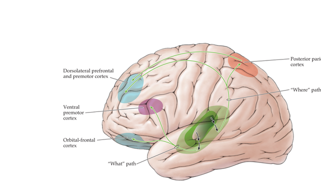
Ongoing efforts

A dynamical systems model is flexible framework to explore

- Statistical inference
- Bayesian tracking
- Multi-scale integration

Considerations/ongoing efforts

- Assumptions of stationarity need to be examined (adaptive coding)
- Translation of memory (short and long) onto attentional feedback signals and its integration in the scheme
- Interplay of bottom-up and top-down feedback



Model implications and validation

- Ongoing tests of model credibility
- Experimental predictions (e.g. how feedback shapes sensory processing)

References

- Chakrabarty, D. & Elhilali, M. "A Gestalt inference model for auditory scene segregation". *PLOS Comp. Bio* 15(1):e1006711, 2019.
- Kaya EM, Huang N, Elhilali M "Neural signature of salience in natural auditory scenes", *under review.*
- Huang N, Elhilali M "Bottom-up and top-down auditory attention", *under review.*

Research support from **National Institute for Aging** under project: **U01AG058532**.